DOCUMENT RESUME

ED 340 760 TM 018 016

AUTHOR Morante, Edward A.
TITLE General Intellectual Skills (GIS) Assessment in New Jersey.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE 91
CONTRACT NCES-75.105
NOTE 45p.; Commissioned paper prepared for a workshop on Assessing Higher Order Thinking & Communication Skills in College Graduates (Washington, DC, November 17-19, 1991), in support of National Education Goal V, Objective 5. For other workshop papers, see TM 018 009-024.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adult Literacy; Cognitive Measurement; *College Students; Communication Skills; Critical Thinking; *Educational Assessment; Higher Education; Problem Solving; Program Evaluation; Remedial Programs; *State Programs; *Student Evaluation; *Testing Programs; *Thinking Skills; Writing Evaluation
IDENTIFIERS *General Intellectual Skills Assessment; National Education Goals 1990; New Jersey College Outcomes Evaluation Program; Performance Based Evaluation

ABSTRACT

College-level assessment efforts in New Jersey are described. For more than a dozen years, students in public colleges and universities in New Jersey have been assessed using a common statewide instrument, the New Jersey College Basic Skills Placement Test. This test assesses reading, writing, and mathematics skills of entering college students as part of an assessment of students and an evaluation of remedial programs. In 1985, New Jersey embarked on a more ambitious program to assess higher education including student learning and development, impact on community and society, and outcomes of faculty research. The new statewide assessment effort, the College Outcomes Evaluation Program, features development of a "sophomore test," the General Intellectual Skills Assessment, aimed at assessing higher order skills of critical thinking, problem solving, quantitative reasoning, and writing. As the nation struggles with questions of assessing the literacy of college students and other adults, New Jersey has already developed a workable system including a reliable and valid performance assessment. Three appendices and one chart provide further information about the New Jersey assessments. A seven-item list of references is included. Reviews by R. L. Larson, M. Scriven, and R. G. Swanson of this paper are provided. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340760

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

GENERAL INTELLECTUAL SKILLS (GIS) ASSESSMENT IN NEW JERSEY

Edward A. Morante
Fall 1991

Abstract. For more than a dozen years, students enrolled in public colleges and universities in New Jersey have been assessed using a common statewide instrument, the New Jersey College Basic Skills Placement Test. The test assesses the basic skills (reading, writing, and mathematics) of entering college students as part of a broader Basic Skills Assessment Program designed to assess students and to evaluate the remedial programs offered at each institution.

In 1985, New Jersey embarked on an even more ambitious program to assess higher education in the state including student learning and development, community/society impact, and the outcomes of faculty research. Building on the experience of the basic skills program, the new statewide assessment effort called College Outcomes Evaluation Program (COEP) featured the development of a "sophomore test" labelled GIS Assessment. This test was aimed at assessing college students' proficiency in the higher order skills of critical thinking, problem solving, quantitative reasoning, and writing.

As the nation struggles with the questions of whether and how to assess the literacy of college students and other adults, New Jersey has already developed a workable system including an innovative, performance assessment (no multiple-choice questions) that is reliable and valid.

This paper describes the assessment efforts in New Jersey with a special focus on the development and implementation of the GIS Assessment. Throughout the paper, questions and problems are raised and then addressed both conceptually and in terms of practical solutions. The very existence of these statewide tests is the best answer for whether such testing is feasible. The paper concludes by raising the question of whether the educators have the will to test and whether the nation can face the test results.

TM018016

BEST COPY AVAILABLE

GENERAL INTELLECTUAL SKILLS (GIS) ASSESSMENT IN NEW JERSEY

Edward A. Morante
Dean, School of Educational Resources,
Research, and Technologies
College of the Desert
Palm Desert, CA 92260

Fall 1991

National Center for Education Statistics
Washington, DC
Contract # 75.105

This paper represents the views of only the author who served as Director of both the Basic Skills Assessment Program (1981-84) and the College Outcomes Evaluation Program (COEP) (1986-91) for the New Jersey Department of Higher Education.

GIS ASSESSMENT IN NEW JERSEY

The state of New Jersey, beginning in 1985, embarked on a comprehensive statewide effort to assess and improve the higher order thinking skills of its college students and graduates. What was developed through this process broke new ground in assessment. The ideas, the process, the program, and the results have implications well beyond the borders of one state. This paper will attempt to describe New Jersey's efforts and relate its implications to a national movement to assess what some call general intellectual skills (GIS) including critical thinking, problem-solving, quantitative reasoning, and communications.

Historical Overview

In the mid-1970's, higher education leaders in New Jersey expressed concern about the academic preparation of many students entering the state's colleges. Following a study by a presidential task force, the Board of Higher Education in 1977 created a statewide Basic Skills Assessment Program. This effort was aimed both at assessing students as they entered college as well as evaluating the effectiveness of remedial programs offered by each institution. At that time, the Board hoped that such a program would self-destruct at the end of five years by successfully eliminating "the basic skills problem." (More detail on the Basic Skills Assessment Program is provided below.)

In the early 1980's, academicians across the country were expressing concern about the quality of the educational enterprise at all levels including higher education. The issuance of "A Nation At Risk" in 1983 had made the public aware of the significant problems facing education. A spate of subsequent reports on higher education called into question the integrity of the curriculum, the outcomes of our endeavors, and the very quality of the college degree. In 1985, the New Jersey Board of Higher Education created the College Outcomes Evaluation Program (COEP) to assess the effectiveness of public higher education in the state. A significant component of this program was a call for a sophomore test.

The development of COEP, including the "sophomore test" are intimately related to the basic skills program. The issues faced, the program developed, and not infrequently the individuals involved in the basic skills effort lay the groundwork for the statewide outcomes assessment program and the development of a second statewide exam.

Consequently, the next section of this paper will provide a summary of New Jersey's Basic Skills Assessment Program (BSAP) including information on the New Jersey College Basic Skills Placement Test and results of evaluating college remedial programs.

The following section will summarize the College Outcomes Evaluation Program (COEP) while the remaining bulk of the paper will focus on the GIS Assessment, the "sophomore test" developed as part of COEP.

BASIC SKILLS ASSESSMENT PROGRAM (BSAP)

In 1977, the New Jersey Board of Higher Education passed a resolution establishing the BSAP (Basic Skills Assessment Program). This effort had two main functions: to assess the basic skills proficiencies of students entering public higher education statewide; and to evaluate the character and effectiveness of the remedial programs at all public colleges and universities in the state. The BSAP was an important component of the Board's two overarching goals: access and excellence.

The 1977 Board resolution also established a Basic Skills Council to oversee the administration of the program and to make policy recommendations to the Board and to the Chancellor of Higher Education. Originally twelve members (now fifteen), the Council is composed largely of faculty and staff from a cross-section of New Jersey's colleges and universities. In addition, a number of committees were appointed by the Council including: Reading and Writing, Mathematics, Tests and Measurements, Assessment, and the Task Force on Thinking.

After several early meetings of the Council, one of its members was elected as the Director. A Basic Skills office was also created in the Department of Higher Education (DHE) and given a support staff. In order to ensure the Council's semi-autonomous nature, including a Director protected from the political machinations of the Department, the Council adopted by-laws which called for the selection of a Director from the ranks of tenured faculty and staff of New Jersey colleges. The director would serve a two-year term on a leave basis and then return to his/her institution. (More recently, the Director has become a permanent DHE staff member with a consequent diminution of independence on the part of the staff.)

The New Jersey College Basic Skills Placement Test

Given its two functions of assessing students and evaluating programs, the Basic Skills Council began its work by focusing on testing students. A survey of the testing practices at each public institution in the state in 1977 indicated a wide diversity across institutions in policies, tests used, and standards. Some colleges had mandatory testing and placement, most had voluntary programs, a few had no testing or remedial programs.

The Council decided that any testing program in New Jersey had to serve two essential functions:

a. assist colleges in placing entering students in appropriate levels of courses, from basic skills to college level; and

b. report to the public on the basic skills proficiencies of all students entering public higher education in the state.

While the first function allowed for a diversity of testing instruments, the second called for a common statewide test since a multiplicity of tests would seriously dilute the meaning of the results. The Council was convinced that any successful effort to impact the educational system in the state, to upgrade educational standards, and to markedly decrease the need for remediation at the college level necessitated a common test. The mixed message sent by different tests and/or different standards would likely confuse the public and have little meaning or effect.

Before turning to the question of what kind of test to use, the Basic Skills Council next defined what they meant by "basic skills". (Basic Skills Council, 1991, p. 1):

> By "basic skills" the Council means the tools of intellectual discourse used in common by participating members of all academic communities. These tools are the language of words and the language of mathematics. Students need these tools to extract information, to exercise and develop the critical faculties of the mind, and to express thoughts clearly and coherently.
>
> Without them, learning is impaired, communication is imprecise, understanding is impossible. A test of "basic skills," therefore, is a test to determine whether an individual has developed the practical working skills of verbal and mathematical literacy needed to take advantage of the learning opportunities that colleges provide.
>
> To define "basic skills" in this way is not to deny the validity of other modes of communication--within the artistic realm of discourse, for instance, the languages of music, motion, image, color, light, and texture express a universe of perceptions, feelings, and emotions which cannot be expressed adequately by words and numbers and logic alone. Nor is the Council's definition of the "basic skills" inimical to the value of diversity. We are, to the contrary, exceedingly sensitive to the differences between colleges: differences in their students; differences in their curricula and pedagogical philosophies; differences in their missions. But in one respect all colleges are identical: their ultimate purpose is to foster learning. The Council asserts unequivocally that the "basic skills" of reading, writing, and mathematics are a prerequisite to learning at the college level. If the possession of these skills is "standardization," we believe that standardization in this sense is good.

After reaching a consensus on a definition of basic skills, a review of tests used across the country convinced the Council to create its own. With the technical assistance of the College Board and Educational Testing Service, the New Jersey College Basic Skills Placement Test (NJCBSPT) was created. The test currently* has five components:

1. Reading Comprehension: A multiple-choice assessment of reading at a level needed for college including inferential reasoning and comprehension, with vocabulary assessed in the context of paragraphs of various lengths and difficulty;

2. Essay: A twenty-minute, holistically scored writing sample graded by selected and trained faculty readers in a common setting using common standards and models (rangefinders). A single topic is carefully chosen each year from over 100 submitted;

3. Sentence Sense: An assessment of writing skills in a multiple-choice format which requires students to understand and apply commonly accepted standardized English in a variety of writing formats;

4. Computation: An assessment, in a multiple-choice format, of elementary arithmetic problems including fractions, decimals, and percents as well as estimation and word problems; and

5. Elementary Algebra: A multiple choice test requiring the student to solve problems commonly taught in a traditional elementary algebra course.

The level of difficulty of the test includes:

Reading and Writing: commonly taught up to tenth or eleventh grade level in high school.

Computation: arithmetic commonly taught before the eighth grade where the most difficult question is a percent problem in the format: "12 is 15% of what number?"

Elementary Algebra: Algebra generally taught in a typical ninth grade where the most difficult problem requires the solving of a simple linear equation with alphabetical functions such as: "ax = c - bx, solve for x."

---

*A sixth component, Logical Relationships, was dropped after three years when analysis revealed that it was too closely related to the reading and writing components of the test.

Reporting Test Results

The NJCBSPT is a criterion-referenced test: a wide cross-section of New Jersey college faculty members identified the basic skills factors that were needed for college and set the proficiency standards for each section of the test. Three levels of proficiency were identified for each of the following: verbal skills (combining the three verbal sections), computation, and elementary algebra. These levels include: Proficient, Lack Proficiency in Some Areas, and Lack Proficiency. In setting these standards, the Council recommended and the Board of Higher Education approved a resolution which called upon all public colleges in the state to set among their multiple criteria for determining need for remediation, cut scores on the NJCBSPT no lower than the middle category, Lack Proficiency in Some Areas.

The first test results of the NJCBSPT were published in 1978 and have been reported publicly every year since. Except for a slight improvement among recent high school graduates, the results have changed little over the years. The results for Fall 1990 entering students are as follows:

In verbal skills,
- 24% Proficient,
- 40% Lack proficiency in some areas, and
- 37% Lack proficiency

In computation,
- 32% Proficient,
- 25% Lack proficiency in some areas, and
- 43% Lack proficiency

In elementary algebra,
- 13% Proficient,
- 29% Lack proficiency in some areas, and
- 58% Lack proficiency

Initially, the results were received with shock and dismay. At the college level, numerous individuals challenged the results, questioning the standards and the test itself. Several studies were carried out on the reliability and validity of the test; faculty on many campuses reviewed the items and the standards. The NJCBSPT was found to be reliable and valid. In addition, there was a broad-based consensus that the standards were appropriate across all levels of higher education: university, four-year state colleges, and community colleges. (Several institutions found it necessary to use an additional test of mathematics (intermediate algebra and pre-calculus) for some of their students).

Remedial Program Evaluation

The second major component of the Basic Skills Assessment Program is the evaluation of each institution's remedial program in reading, writing, and mathematics. Since 1980, common definitions and reporting formats have been used. Students who need and complete remedial programs are compared to students who did not need remediation in a particular basic skill area. Two year cohort analysis is carried out using several outcome variables: retention rates, pre- and post-testing, GPA, passing rates in subsequent college courses, and academic success rates (combining retention with GPA). Results indicate great diversity in both types of programs and program effectiveness from college to college. Across the state, students who complete remediation are retained at comparable, if not slightly higher rates, but achieve somewhat lower grades than students who did not need remediation. (Morante, 1986)

More recently, the Basic Skills Council has recommended, and the Board of Higher Education has approved, standards which each program is expected to achieve. For example, 90% of students who complete remediation are expected to demonstrate proficiency on the post-test, an alternate form of the NJCBSPT. (Ninety percent was used instead of 100% to account for the error variance of the test.

Accomplishments of the Basic Skills Program

Probably the most significant accomplishment of the BSAP is the raising of standards in New Jersey at both the K-12 level and the college level. At the time of the beginning of the BSAP in 1977, New Jersey required its high school students to pass a test as part of the requirements for receiving a diploma. Unfortunately, the standards for this basic skills test was set at about the sixth grade level. In 1981, the Commissioner of Education declared a success story: basic skills standards had been achieved. He announced this conclusion despite continuing declines in SAT scores and the stark results of the NJCBSPT. The Chancellor of Higher Education publicly chastised this announcement as false optimism and detrimental to quality education. In the midst of a gubernatorial election, Tom Kean sided with the Chancellor and promised to raise standards and to fire the commissioner if elected. He was, he did fire the commissioner, and two weeks after a new commissioner was appointed, the announcement was made to end the easier test and to create a new, more difficult high school proficiency exam. Several years after that, the Governor announced the phase-out of the ninth grade test and the development of a more comprehensive eleventh grade test. In 1995, eighteen years after the basic skills program began, New Jersey will require students who seek a high school diploma to pass a basic skills test that is comparable to the standards called for on the NJCBSPT and by the Basic Skills Council.

The basic skills program has also had a profound effect on higher education in the state. These include:

1. A significant increase in standards and expectations of basic skills proficiency needed for college level courses. This is most evident in the raising of the cut-scores at many colleges, sometimes markedly. These increases in cut-scores resulted in placing many more students in remedial courses rather than permitting them to enter college courses without adequate preparation.

2. Expansion and much greater acceptance of the need to provide comprehensive developmental education programs at all colleges and universities. This was critical in demonstrating that both access and quality were achievable goals.

3. Increased communication among faculty members across and within institutions. Many institutions formed committees to discuss changes in curriculum, teaching, and services provided to improve students' basic skills.

4. Creation and/or expansion of research and follow-up of students for institutional decision-making and improvement.

COLLEGE OUTCOMES EVALUATION PROGRAM (COEP)

The New Jersey Board of Higher Education passed a resolution in June, 1985, calling for a comprehensive assessment program of public higher education in the state. The resolution called for assessment in such areas as general education and the major, retention and graduation rates, and community and society outcomes. An important component of the endeavor was a call for a sophomore test: "That the evaluation system shall include an assessment of students' learning through the administration of a test battery that measures proficiency in writing, quantitative reasoning, critical thinking and any other areas appropriate for the evaluation of general college-level academic proficiencies." (BHE, 1985)

Following the model of the Basic Skills Program, the Board appointed a broad-based Advisory Committee to recommend the details of the assessment program. This committee, chaired by the powerful president of the state's medical university, consisted of twenty-four individuals representing higher education (public and private), K-12 education, state agencies, businesses, and the public, including students. Following its first meeting, the committee appointed four subcommittees, with additional representatives of higher education to address four key areas: student learning, student development, community/society impact, and faculty research. Subsequently, the former director of the Basic Skills Assessment Program was appointed Director of COEP. An

office was created in the Department of Higher Education which at its high point consisted of five professionals and two secretaries.

Many meetings were held from 1985 to 1987. In addition to regular committee meetings and numerous campus meetings, COEP sponsored several statewide conferences to elicit ideas and to provide feedback to draft reports. A major boost was given for the program in 1986 when then Governor Thomas Kean proclaimed at the largest conference on higher education ever held in New Jersey: "I support the COEP effort as promulgated by the Board of Higher Education."

The Board formally adopted all of the recommendations made by the COEP Advisory Committee in October, 1987. Chart I provides an overview of the components of this comprehensive statewide assessment program. (The Report to the New Jersey Board of Higher Education, 1987, p. vi-vii)

CHART I ABOUT HERE

At this same time, the Board also appointed a standing COEP Council to oversee the program and called upon the Council to provide periodic reports assessing higher education in New Jersey. (See Appendix A for a listing of various reports developed or contracted for by COEP.)

During the period of the creation and development of COEP, New Jersey was in the midst of an economic boom including strong support, both financial and political for higher education. The Governor (Kean) and the Chancellor (Hollander) were considered national leaders of education. Large sums of money were distributed in the form of grants (especially "Challenge" grants) to spur local educational improvement. COEP and its emphasis on improvement and accountability were encouraged by an influx of funds, by state leadership, and by efforts, including legislation, to encourage autonomy (at the state colleges). The message was, "We'll give you more flexibility to run the day-to-day operations of your institution and increased funding for innovative plans for improvement, if you agree to be held more accountable for demonstrating the outcomes of your efforts."

Toward the end of the decade, the economy slowed and the budget for COEP was cut. Several state college presidents and the leadership of the state college faculty union voiced serious concerns about COEP, especially the GIS Assessment (described below). The governor and the Chancellor strongly supported the

CHART I
COEP VARIABLES

I Outcomes Clusters/Variables	II Type/Source Of Data	III Collection/Reporting Frequency
1. General Intellectual Skills-students' ability to find, use, & present information/data; skills in analysis, problem solving, critical thinking, quantitative reasoning, verbal abilities	Statewide assessment; test *samples* of students for *institutional* assessment	Periodic public reporting
2. General Education — defined partly as (a) ability to understand & apply *modes of inquiry*, (b) appreciate & confront enduring aspects of *human condition*, variety of responses to human issues & problems, and fashion reasoned ethical responses	Locally developed assessment defined partly in accordance with statewide definition	Periodic public reporting
3. Major field of study — defined in terms of objectives/outcomes chosen by faculty in each program/department	Program-level assessment	Part of ongoing 5-year evaluation and reporting cycle for all programs
4. Indirect indicators of student learning a. Retention rates b. Grade point averages c. Credit completion rates d. Program completion rates e. Licensure exams	Common definitions for (a) to (e)	Periodic public reporting for (a) to (e)
f. Students on academic probation/dismissed g. Reasons for withdrawal h. Graduate/professional school exams	Local definitions for (f) to (h)	Periodic internal reporting for (f) to (h)
5. Student involvement and satisfaction a. Enrolled students' involvement & off-campus activities b. Enrolled students' satisfaction	Locally defined	Periodic internal reporting
6. Students' personal development a. self-awareness b. values c. interpersonal relationships d. leadership	Locally defined	Periodic internal reporting
7. Community/Society Impact a. Human Resource Development: training/job related programs offered; projections on labor force needs; employer needs and perceptions re quality of students	Institutions collect/analyze data, with common definitions and designs	Periodic public reporting
b. Access: percent of target subgroup members admitted as students &/or receiving services, compared to demographics of region/community	Admission data from SURE; institutional data re participation in programs/services; surveys of needs assessment and perceptions of access	Periodic public reporting
c. Economic Impact: e.g., expenditures; economic contribution by institutional employees, students; data on costs of city services; taxes	Institutions compile & analyze data; report to COEP	Periodic public reporting
d. Local priorities	Locally defined	Periodic public reporting
e. Post-Collegiate Activities: e.g., further education, employment, satisfaction, and community activities	Institutions collect/analyze data with common definitions and designs	Periodic public reporting
8. Research, Scholarship, & Creative Expression (e.g., dissemination of knowledge/methods/ new discoveries to students, peers, business, & industry)	Defined in consultation with institutions; possible combination of statewide and locally selected outcomes	Periodic public reporting

assessment program and found the funds to maintain the program. (At the Chancellor's request, Governor Kean allocated $150,000 from a contingency fund specifically to continue the administration of the GIS Assessment.)

The election of a new governor in New Jersey in 1989 sealed the fate of COEP. In rapid succession, the Commissioner of Education was fired and the Chancellor of Higher Education resigned (after the new governor allegedly refused to see him or accept his telephone calls). Within months of his selection, the new Chancellor cut the staff of COEP by 60 percent by transferring them to other areas of the Department. When the COEP Council objected and threatened to resign, the Chancellor backed off until July 1, 1991 when he ended COEP and terminated the Director.

In its six years, COEP:

1. developed arguably the most comprehensive statewide assessment program in the nation - a model that has positively impacted other states and many institutions;

2. refocused higher education in New Jersey to include an emphasis on outcomes in addition to the historic examination of inputs and processes;

3. created the first statewide assessment of higher order general intellectual skills in the country;

4. broke new ground in testing technology in implementing a reliable and valid instrument which simulates actual academic performance without reliance on traditional multiple choice questions;

5. broadened the definition of research to include scholarship, creative expression, and teaching activities;

6. reported on the first cohort longitudinal analysis of retention and graduation at every public institution in New Jersey as well as transfer between institutions;

7. became the first state in the country to include an operational definition for assessing access, retention, and graduation rates for minority students in public higher education and recommended goals for institutions to achieve;

8. included for the first time in the country a statewide focus on assessment of institutional impact on the local community, including public service activities and economic impact;

9. developed the first statewide survey in New Jersey for the assessment of a common core of post-collegiate activities of former students at each of New Jersey's public institutions;

10. organized the largest conference on New Jersey higher education in the history of the state.

11. fostered assessment and frequently redefinition of general education across the state at both public and private institutions;

12. impacted the redirection of the Middle States' accreditation process toward focussing on outcomes;

13. significantly increased examination of the goals, objectives, and outcomes of most of the majors in New Jersey public higher education;

14. actively involved hundreds of New Jersey faculty and staff members in planning for, creating, and implementing a statewide assessment program;

15. directly affected most of the state's institutions to examine their missions, strategic planning, programs and impacts.

GIS ASSESSMENT: THE "SOPHOMORE TEST"

Rationale

Far and away, the most controversial aspect of COEP was the Board's call for a sophomore test. Over time, the basic skills test had become an accepted aspect of higher education in New Jersey, but the NJCBSPT was not perceived as much of a threat to college faculty or administrators. It was not the fault of higher education if students entered inadequately prepared, especially at the community colleges which espoused open assess and the opportunity to develop these and higher order skills. But a test that assesses students after completing a sizeable portion of their college education was perceived as a measure of the effectiveness of the college education they received. This was scary to many in the state's colleges and universities.

In addition, the principal model of a sophomore test in the country in 1985 was Florida's CLAST program. In that state, students are required to pass a "sophomore test" in order to receive an associate degree or to be permitted to continue their education into the junior year. It was widely assumed in New Jersey, especially in the year following the Board of Higher Education resolution, that this form of "gateway test" (sometimes referred to as "rising-junior exam") would become the testing program in New Jersey.

There is little doubt that the fear of a sophomore test was directly related to a fear of being held accountable. Board of Higher Education members publicly proclaimed that accountability as

well as using the information to improve student learning, were the reasons for implementing an assessment program including a sophomore test common across all public colleges and universities.

However, a series of additional concerns beyond accountability were expressed. Many of the following issues were raised in the context of assuming that the Florida model of sophomore test would be the prototype of the program in New Jersey:

1. A gateway test is harmful to students by preventing them from continuing their education;

2. Traditionally underrepresented students especially minority students, would be most seriously impacted;

3. A gateway test would place the burden of responsibility on students rather then on the faculty or the institution to seek improvement in teaching and learning;

4. A basic skills test at the sophomore level will imply that this is all that is expected of higher education and will lead to lowering standards;

5. A strength of the American higher education system is its diversity and any common test would undermine that diversity and result in weakening higher education;

6. In order to achieve politically acceptable passing rates, the test would have to cater to the lowest common denominator and thus undermine standards and quality institutions;

7. No multiple-choice test could adequately measure the college-level skills expected of students (this point was frequently supported and bolstered by any negative criticism of any standardized test ever developed for any use);

8. Students take too many tests now or faculty already give tests in their courses;

9. A statewide test drives the curriculum or what is measured is what is important.

Toward the end of 1985, the committee designated responsible for addressing these and similar concerns, the Student Learning Outcomes (SLO) sub-committee, began its deliberations. The committee membership included several individuals who had been active leaders in the Basic Skills program including the chair, a faculty member who had formerly chaired the state's Basic Skills Council. The experiences from the earlier developed program were invaluable in tackling the charge given to it: to develop a statewide system to assess student learning. The committee was not intimidated by the external concerns and hostility. A similar

climate had existed at the early stages of the basic skills program and had largely dissipated over time. While some committee members were openly opposed to any program, the overwhelming majority were willing to attempt to create a viable system. It was also evident that almost every member of the committee had a serious concern about student learning and that any recommendation would have to include this concern as well as focus on ways to improve student learning.

Early in its deliberations, the committee agreed to postpone any discussion of the methodology of assessment until a conceptual framework could be formulated. They also agreed that no discussion of a statewide test would take place unless some agreement could be reached on whether there existed any commonality across institutions. The first major breakthrough occurred when it was agreed that two factors were universally accepted (some felt: or should be universally accepted) by all faculty members as required for all students: writing and critical thinking. The Basic Skills program again served as a backdrop for the ensuing discussions since it was commonly accepted from the experience of the basic skills test (NJCBSPT) that at least writing could be successfully assessed statewide.

Over the next two years, discussions within the committee and with external groups, as well as reviews of the literature, gradually led the SLO committee to reach the following conclusions:

1. Diversity was a strength that should be maintained; however, commonality across institutions was also clearly evident. Each institution's uniqueness was related more to its own special combination of factors (students, faculty, facilities, resources, mission, etc.) than to the uniqueness of each individual aspect. For example, each community college (at least in New Jersey) had an open-door policy for admissions, each served a wide diversity of students, each had both transfer and terminal programs, each required some combination of general education courses, and so on. While no community college was exactly the same as any other college in the state, all of them shared qualities that were seen in other institutions. In mathematical terms, the sets were not mutually exclusive.

2. A conceptual formulation was developed that differentiated between general education and the major as the two essential components of a higher education curriculum. Further, the committee separated what they ultimately called "general intellectual skills" from the content of general education. These general intellectual skills (GIS) were the equivalent of the traditional skills of critical thinking, problem solving, quantitative reasoning, and communications (both oral and writing). Specifically, the SLO committee made the following definitions. (See Appendix B for the

committee's more comprehensive definition.)

> a. Accumulate and Examine Information (Gathering Information) -- including the skills necessary to: determine the kinds of information needed for a given task; construct and implement systematic search procedures using both traditional and computerized methods; discard or retain information based on an initial screening for relevance and credibility; and develop abstract concepts appropriate to the task at hand for initially ordering the information which is retained.
>
> b. Reconfigure, Think About, and Draw Conclusions from Information (Analyzing Information) -- including the skills necessary to: evaluate the interpretations presented by others in terms of their assumptions, logical inferences, and empirical evidence; reconfigure information in ways that suggest a range of alternative interpretations and evaluate their relative merits; construct hypothesis that logically extend thought from areas in which information is already available into areas where it is not; specify the additional information which might confirm or disconfirm those hypotheses; and draw conclusions based on all of the above.
>
> c. Present Information (Presenting Information) -- including the skills necessary to express one's own ideas in written, oral, and graphic forms which will be intelligible and persuasive to a variety of audiences. (COEP Advisory Committee, 1987, p. 10)

As a result of test development, a fourth area was added for scoring purposes: Quantitative Analysis which "replicates analyzing information but concentrates on problems requiring quantitative reasoning and calculations." (COEP Council, 1990)

3. A statewide test in general intellectual skills was desirable and could have a positive impact on improving student learning. As the Basic Skills Council had concluded almost ten years before, a statewide test external to each institution would be the best method of impacting all of the institutions, would provide the best and fairest method of accountability, would be the most likely to ensure that these skills are taught (agreeing with the statement that what is measured is important). Further, a common statewide test would be the most economical and efficient means of test development because it would pool limited resources and put the funding support of the state behind the assessment effort. Without an external assessment, it was feared that many institutions would set standards more aligned with the proficiencies of the student body than against some common criteria or standard. The committee felt that many colleges had developed standards for awarding the degree in this way, a phenomenon that has occurred with the high school diploma.

4. Any test that would be developed should emphasize institutional responsibility rather than place the burden of improvement on the students. Thus, the concept of a gateway test was rejected; the committee accepted the consequences of their decision in recognizing that motivating students to perform their best on any test would need to be addressed;

5. A sophomore test should assess higher order skills beyond the basic skills measured by the NJCBSPT and that the test should, as much as possible, model the academic skills expected by faculty. The committee wanted a test that, if possible, avoided multiple-choice questions and instead required students to demonstrate directly the skills expected. A major breakthrough in achieving consensus on this point was reached when the committee examined the "assessment center" approach used in industry, especially that developed in the Bell Telephone system. (In fact, the tasks later developed in the GIS Assessment are academic cousins of the "in-basket," a technique commonly used by industry to assess individuals for hiring or promotion into management positions.)

6. The test, in addition to being reliable, valid, and unbiased, would need to address the breadth of standards expected of students at different colleges. The test should be seen by faculty as being both "reasonable" and "challenging" rather than "too easy" in terms of requirements and standards. An important premise underlying this statement was an agreement that there is such a thing as a minimal standard across all colleges, (otherwise, why use the term "college"?).

In reaching this consensus, the committee members extensively discussed the wide diversity of academic preparation students bring to different colleges. They questioned whether, given this diversity, it was reasonable to expect a minimal standard across all institutions in the state. Again, the existence of New Jersey's statewide basic skills program played an important role in their conclusion that such a standard was appropriate. The BSAP required that students who entered college lacking appropriate levels of basic skills would need to acquire those skills before entering college level courses. This process of remediation, if effective, would greatly decrease the entering diversity. Each institution has the responsibility to both teach and ensure learning of basic skills to those it accepted and who needed such remediation or to dismiss the students who were unable or unwilling to learn (the emphasis on retention provided a counter-balance to mass dismissals). Of course, this is also a question of standards. If a college awards credit for courses students complete, that institution must be held accountable for ensuring that students are learning. If general intellectual skills are or should be an integral part of the curriculum, in all or nearly all courses, a college should be responsible for its students achieving some

minimal level of proficiency in these general intellectual skills after they complete these courses. The very essence of awarding credit (and by implication a degree) assumes this institution has taken responsibility for student learning and proficiency.

The COEP Advisory Council accepted the report of the SLO committee and incorporated its recommendations in its October 1987 report to the Board of Higher Education. The next step was to actually develop a test.

Development Of An Instrument

Work on the development of a sophomore test began immediately after the Board adopted the recommendation of the COEP Advisory Committee. A contract was awarded to Educational Testing Service (ETS) to provide technical assistance and management to the test development. The Advisory Committee became a Council, a newly constituted GIS Assessment Committee of faculty and staff was formed, a Task* Writer's Sub-committee was constructed, and a former state college dean (and member of the SLO Committee) with direct responsibility for the test development was added to the staff of the COEP Office within the Department of Higher Education.

The task writers were all New Jersey college faculty members respected on their campuses for quality teaching. More than sixty tasks were written covering three major components of general education: the arts and humanities, the natural sciences, and the social sciences. Many of these tasks were tried out in actual classrooms to elicit feedback from students and faculty. Subsequently, 27 tasks were selected for the first pilot administered to 2,663 students at 16 institutions in New Jersey during Fall, 1988. The results of the first pilot demonstrated the feasibility of the concept of the GIS Assessment, including the ability to select and train faculty to appropriately score the student responses.

The test development process also included writing detailed scoring guides for each task and the development of a core scoring system (where a "4" on a 6-point scale was set as appropriate proficiency for students completing the equivalent of sophomore year). In addition, there was preparation of a procedures manual and training for proctors and scorers. Review and feedback were integral parts of test development with input from hundreds of faculty members (including committee members, faculty readers, faculty proctors and reviewers, and a special validity panel), suggestions from national consultants, and critique by DHE staff.

---

*A "task" is a series of materials, questions, and problems presented to each student to assess general intellectual skills. Several examples are given later in this paper and in Appendix C.

In Spring 1989, a second pilot test was carried out. Sixteen revised tasks were administered to 2,201 students from 12 New Jersey institutions. The results of both pilots indicated that the GIS Assessment was ready for operation. In October, 1989, ETS issued its final report on the test development stating:

> •The materials developed for the assessment of general intellectual skills, especially the extended tasks,* are valid and innovative measures of certain college level intellectual skills.
>
> •The extended tasks and scoring processes "worked," that is, students could respond and readers could score them reliably. The GIS Assessment is therefore, an appropriate measure of the skills its seeks to assess ...(ETS, 1989, p. 2)

In December 1989, the Board of Higher Education endorsed the Chancellor's recommendation to implement the GIS Assessment beginning with sophomores enrolled in public college and universities in the Spring of 1990.

What is the GIS Assessment?

The GIS Assessment consists of 14 separate tasks (4 additional tasks were piloted in 1991 and, after some revisions, could be added to the pool of available tasks). To adequately cover the domain of skills, seven tasks, balanced for difficulty and content, have been used in each of the two statewide administrations carried out at all 31 public two- and four-year colleges in New Jersey in 1990 and 1991. (The results for 1990 are presented below.)

Each student takes only one task and is allotted 75 minutes to complete the assignment. The tasks are administered randomly to the students taking the test. No attempt is made to relate the content of the tasks to the student's major or courses completed. The emphasis is on assessing the underlying general intellectual skills needed by all students regardless of major or institution. The results are produced only for groups of students (e.g. institutions) by summing the skills assessed over all seven tasks. Stated differently, each student takes only a portion (i.e., one-seventh) of the GIS Assessment. This permits a broad range of content to be included in assessing general intellectual skills without burdening each student with an overly lengthy assessment. This procedure also precludes using the GIS Assessment as a gateway device -- an important consideration in New Jersey.

---

*Multiple choice items were included in the test development; however, the data indicated that they added little to and, in some cases, lowered the validity and reliability of the test.

In completing one of the tasks for the GIS Assessment, each student receives a packet of materials in a sealed envelope. Following specific written instructions, the student is requested to read the materials and respond to a series of questions. All of the tasks require short written responses and one extended essay; calculations, map reading, and drawing or graphing may be required according to the individual tasks. Generally, the tasks begin with easier questions and become more difficult at the end. The essay is almost always at the end of the task. Two examples of tasks are given below; Appendix C contains summaries of several others.

The Plague. In this task, the student is asked to simulate the drafting of a research paper. Each student who takes this task is presented with a series of twenty 5 x 8 cards containing typed notes as if someone (another student) had gone to the library to gather information about the plaque that ravaged Europe in the middle ages. The student is first asked to organize the cards into major headings, then to answer questions about the material on the cards. For example, one question asks the student to summarize in one or two sentences a lengthy amount of material. Another question calls for the student to compare and contrast several cards. Finally the student is requested to draft a 300-500 word essay using the material appropriately as presented on the cards. Accurate attribution of the correct author (by card numbers) is also expected for the essay.

Cezanne. This task, which might more appropriately be called "Cubism," presents the student with three color postcards of paintings by three artists as well as an essay on the topic. The students are requested to study and analyze the paintings, compare and contrast them, and conclude with an essay summarizing the development of Cubism through the work of these artists.

This task, in particular, presented a special challenge and an opportunity to both teach and assess. Feedback from several sources during the pilot stage of test development confirmed the preconception that few students have taken a course in art history and that many are grossly ignorant of (and usually uninterested in) the topic. The task writers and the GIS Advisory Committee decided not to succumb to the easiest path which would be to drop this task rather than expect students to demonstrate general intellectual skills on a topic few knew much about. In fact, this task on art history probably more than any other task (although some questions on other tasks which required mathematical calculations also faced similar concerns), confirmed the notion that the GIS Assessment could be a teaching device to be used to improve teaching and learning, as well as an assessment instrument. After the first pilot, Cezanne was revised in such a way as to better introduce students to the topic. The materials were presented which requested students to give their perceptions about the paintings and to draw certain figures directly on black and white outlines presented for two of the paintings. Several geometric shapes were

included in one of the paintings to help the student begin the process. The idea was to help the student become interested in the task by requesting him/her to actively participate in exploring the structure of the paintings in a more direct way than merely asking for response to questions.

Scoring the GIS Assessment

In creating the GIS Assessment, especially the use of tasks instead of multiple-choice items, the developers realized that the scoring process would be much more difficult, time-consuming, and expensive. The use of representative sampling (described below) instead of attempting to assess an entire institution or class of an institution markedly decreased the cost. In addition, the experiences developed through the scoring of the essay portion of the Basic Skills Test and further honed in the two pilot tests, gave confidence that scoring the GIS Assessment was feasible. The availability of a core of experienced essay readers in New Jersey also aided the scoring process. And, finally, the COEP committees, especially the original SLO committee and its successor, the GIS Advisory Committee, were strongly convinced that the use of tasks that simulated what students were expected to do in the classroom was essential for the validity of the instrument. With the cost significantly reduced by using sampling, the special requirements needed to score the GIS Assessment were worth the effort.

Detailed scoring guides were a concomitant development of the tasks themselves and they followed a similar process of writing, piloting, and revision. Experience taught us that it was essential to write and revise the scoring guides only in conjunction with both the task and how students responded to the questions in each task. One scoring guide has been written for each task used; in addition, a generic scoring guide has been written for holistic scoring of the writing component of the essays. In fact, the revisions of the scoring guides (and to some extent the tasks as well) are dependent upon the questions, concerns, and ambiguities raised by the faculty readers themselves in the scoring process. The scoring process also demonstrated the ability of the overwhelming majority of faculty members from different backgrounds to reach consensus on standards, on criteria for scoring, and on resolution of ambiguities.

Each question, including each essay, was scored for content on a six point scale separately by two different faculty members using the scoring guides. (Several less difficult questions had only a four point scale.) In addition, the essay questions were scored holistically by two additional faculty members. Differences in ratings were adjudicated by a third reader, a table leader. All of the readers were selected and trained New Jersey faculty members. All of the readings were carried out in central locations. (ETS was used for the pilots while the test administrations were scored

in the building housing the Department of Higher Education.) The training conducted at the scoring sites used actual student responses as "rangefinders" for training (models of each point on the scoring scale). Readers were divided into small groups and led by a "table leader." Each room (two were used in the scoring of each test administration) was led by a "chief reader."

Representativeness and Student Motivation

In the two administrations of the GIS Assessment in 1990 and 1991, each public college and university was asked (in fact mandated by the Board of Higher Education) to select a representative sample of 200-300 sophomores enrolled in the Spring semester. (A sophomore was defined as a student who completed between 45 and 70 college-level credits, including transfer credits.) In this regard, each college was held accountable both for courses taught at its institution and for transfer credits granted for courses completed at other institutions. The sample could be selected randomly (using commonly accepted procedures) and the selected students tested outside of class. A second procedure was also allowed; this involved selecting a cross-section of classes which enrolled sizeable numbers of sophomores and testing all of the students in those classes during class time. The latter method was chosen by a number of colleges since this procedure was more easily accomplished and produced larger (sometimes much larger) numbers of students, not all of whom however fit the definition of "sophomore." Still other institutions invited all of their defined sophomores to be tested; this, too, tended to produce large numbers of assessed students (almost all of whom were sophomores).

The representativeness of the sophomores tested at each institution was calculated by comparing the students who were tested to the full population of sophomores enrolled at each institution. A number of variables including demographics, basic skills and SAT (where available) scores at entry, GPA, and number of credits completed were used in the comparison. (This was accomplished using the Department of Higher Education's Student Unit Record Enrollment or SURE system, a computerized system which includes all students enrolled in all public colleges and universities in New Jersey.) The representativeness was calculated (chi square) for each variable for each institution. This permitted the GIS Assessment results to be adjusted to account for reasonable differences in representativeness. (In actuality, only mean grade point average and Total English, a composite score of verbal skills derived from the NJCBSPT, were meaningfully correlated with scores on the GIS Assessment; SAT scores was also related to performance on the GIS Assessment for the universities.)

While the issue of representativeness could be reasonably measured and accounted for, addressing student motivation to

perform well was more difficult. As indicated above, the SLO (Student Learning Outcomes) Committee originally accepted student motivation as a concern as part of the price to pay for avoiding a totally unacceptable alternative, a gateway test (which obviously all students would need to take and would have direct consequences for each student. The unfair burden on the students and the cost of such a program were the reasons for their rejection of this option.) The SLO Committee was convinced that procedures could be developed so that the overwhelming majority of students could be motivated to take the GIS Assessment seriously.

Student motivation is a complex issue and requires multiple approaches. First, it is important to differentiate between motivating students to come in to take the test versus motivating them to take the test seriously once in the testing situation. In New Jersey, mostly based on the experiences in other states, we predicted that most of the difficulty of motivating students was in getting them to show up for the test. The two statewide test administrations confirmed this expectation. Based on these experiences, the following conclusions seem reasonable.

1. Using intact classrooms is the easiest way to ensure large numbers of students will be tested. This process also tends to produce a good cross-section of students including weaker students. Ensuring representativeness can be handled by the process described above (HEGIS/IPEDS data might be an alternative). However, overtesting beyond the sample size desired, is needed since students simply do not enroll in courses following historic notions of what a "sophomore" is. In New Jersey, the responses of these non-sophomores (mostly freshmen) were included in the scoring and were reported to the institutions, but were not included in determining proficiencies at the institution, sector, or state level.

2. Inviting many more students to volunteer to take the test than is needed for a sample can also produce large numbers of students, although there is a tendency for the better students to show up. However, while the sample may be skewed, students with the full range of proficiencies do come in for the test, usually permitting appropriate statistical adjustments to be carried out. It is not clear what motivates weaker students to agree to be tested. Three hypothesis are offered:

   a. they don't see themselves as weak since many students overestimate their proficiencies;

   b. many students are truly interested in feedback on their performance in college using a "state instrument" that goes beyond their grades at one institution; and

   c. external incentives play a role (e.g., a reward of some kind) in encouraging participation.

3. The messages given to students, especially by faculty members, are frequently crucial for soliciting cooperation. A faculty committed to the testing program will inevidently produce a motivated student body. The reverse is more ambiguous and depends on how vigorously the faculty members are in opposing the test as well as on how the testing is carried out. In all probability, each campus will have a distribution of feelings among the faculty. The leadership of the president and the academic vice-president can make or break such a testing program.

4. Explaining to students the purpose of the test and providing them with feedback on the results can go a long way in soliciting student cooperation (and motivation).

5. Providing external incentives to students can assist motivation but such extrinsic rewards are usually not sufficient. In the two statewide administrations in New Jersey, the institutions which paid students (there were few) to take the GIS Assessment did <u>not</u> solicit a better turnout than many institutions who used other methods. In fact, probably the single best motivation for inducing students to volunteer to take the test (i.e., outside of using intact classes) was a promise by the president to send to each student who performed well on the GIS Assessment, a letter that could be used as a reference in future career or job opportunities.

James Madison University in Virginia has set aside a day for assessment in their academic calendar. Creating an environment which includes assessment as a normal part of academic and campus life has proven to be successful on that campus. (Alverno College is also a prime example of integrating assessment and teaching, but in a very different way.) One community college in New Jersey administered the GIS Assessment at the same time as other assessment instruments including a student satisfaction survey and an assessment of general education and writing. This served to be generally successful in reaching a large number of students. In these examples, issues of equity (everyone participates), commitment (assessment is an important and integral part of teaching and learning), and improvement (using the results to improve the institution as well as the students) are all contributing factors to motivating students.

The testing experiences gained in the GIS Assessment confirmed the prediction that motivating students to come for the testing was more difficult than motivating students to perform well at the time of testing. Nevertheless, ensuring student motivation at the test site was also important. Several factors were used to accomplish this. Perhaps the most significant way was to make the GIS Assessment intrinsically interesting and challenging for the students. Student feedback was included in the test development to ensure intrinsic student motivation; student perceptions in taking the test were also included in the interpretation of the GIS

Assessment results.

Other factors used to motivate students included:

1. Written and oral messages to students explaining the purposes of the test;

2. feedback on performance;

3. extrinsic incentives for quality performance; and

4. partial reliance on the internal desire of most people to be competitive and/or to try their best, especially in a college setting.

GIS Assessment Results

The GIS Assessment has been administered twice as a statewide test. The results of the Spring 1990 administration were presented to the Board of Higher Education in July 1990. (COEP Council, 1990) That first year, 4,683 students from 28 public institutions took the test; of these, 3,701 were sophomores (45-70 credits completed) or about 12% of the total population of sophomores enrolled at these institutions. Statewide, these students were a generally representative cross-section of sophomores. The results were presented in terms of three levels of proficiency:

Demonstrated Proficiency: these students achieved the level of proficiency expected of a student completing the equivalent of two years of college work;

Somewhat Proficient: these students have achieved some proficiency but not at a level expected of a student completing the equivalent of two years of college work;

Did Not Demonstrate Proficiency: these students were clearly below the level of proficiency expected of a student completing two years of a college education.

The statewide results were as follows:

	Demonstrated Proficiency	Somewhat Proficient	Did Not Demonstrate Proficiency
Gathering Information	58%	27%	15%
Analyzing Information	44%	41%	15%
Quantitative Analysis	33%	38%	29%
Presenting Information	51%	26%	23%

Detailed information on both individual and institutional performance has been sent to each institution, but not published.

The scoring process worked well yielding an interreader reliability coefficient of .82 across all ratings. Ninety-five percent of the students who completed the GIS Assessment reported that they had made at least some effort in completing the test. Colleges reported that they had much more difficulty in motivating students to come in for the test than to take it seriously once in the testing situation.

The first administration of the GIS Assessment was not without trouble, however. The faculty union at the state colleges (little or no faculty resistance was evident at the community colleges or at the universities) organized a boycott of the test which significantly reduced the number of students tested at most of these institutions. At several colleges, including a few community colleges, lack of positive leadership by top administration also reduced the number of students tested.

In 1991, there was no organized boycott and many campuses reported sizeable increases in the number of students tested. More than 6,000 students were tested, but the results have not yet been published.

Conclusion from the GIS Assessment

The most important conclusion from the New Jersey experiences is that statewide testing, and by implication, nationwide testing is possible not only at the level of basic skills but at a level of higher order skills as well. It is feasible to reach consensus on definitions of skills needed for a college education and by direct implication for an educated citizenry. It is feasible to construct reliable and valid measures of these skills.

What is much more difficult to accomplish is the will to try. Many faculty and administrators will oppose any external evaluation -- especially one that has common definitions and standards. The basic skills program in New Jersey demonstrated, however, that over time, given careful consideration and broad-based input, a reasonably conceived and implemented program can be accepted and can work. The COEP effort, especially the GIS Assessment, demonstrated that such a concept can become operational given sufficient resources and leadership. The strong opposition by some administrators, especially college presidents, and at least one faculty union (there are three major organizations in New Jersey) have implications for implementation. Change is not easy in higher education, but possible. Unfortunately, the political climate of a state can change radically by an election. We may never know whether the GIS Assessment would have had the desired effect.

There is evidence, however, that at least at some colleges faculty and administrators as a result of COEP and the GIS Assessment were looking at ways to improve their student learning (and test scores) by reviewing curriculum. Several campuses in recent years had begun to focus on emphasizing more writing in many courses as well as including more essay-type questions on final exams. Some faculty were exploring ways of using GIS Assessment-like tasks in their teaching. One faculty member quit the Task Writers subcommittee and rewrote the syllabus for freshmen physics at his college to incorporate what he had learned from the experience of being a task writer. A cadre of faculty was beginning to emerge on the teaching implications of the GIS Assessment.

The future of the GIS Assessment at this time is very uncertain. Perhaps its greatest accomplishment was in demonstrating its own feasibility. Perhaps its greatest liability was in demonstrating the results.

APPENDIX A
COEP REPORTS

Report to the New Jersey Board of Higher Education from the Advisory Committee to the College Outcomes Evaluation Program (1987)

Appendices to the Report to the New Jersey Board of Higher Education from the Advisory Committee to the College Outcomes Evaluation Program (1987)

Personal Development and the College Student Experience: A Review of the Literature

Procedures Manual for the Assessment of Community/Society Impact at New Jersey Institutions of Higher Education

The Academic Performance of Students Who Began at Eight State Colleges in Fall 1986 (1988)

Planning for Assessment: Mission Statement, Goals, and Objectives

The Assessment of Student Development Outcomes: A Review and Critique of Assessment Instruments

A Report on the Development of the General Intellectual Skills Assessment (ETS, 1989)

Appendices to a Report on the Development of the General Intellectual Skills Assessment (ETS, 1989)

A Report of the College Outcomes Evaluation Program (COEP) Council to the Board and Department of Higher Education on the Development and Implementation of a Statewide Test to Assess the General Intellectual Skills of College Students (1989)

1991 Administration Procedures for the General Intellectual Skills (GIS) Assessment

General Intellectual Skills (GIS): Information Bulletin

Handbook for Calculating Short-Term Economic Impact at New Jersey's Institutions of Higher Education

Assessing Higher Education's Outcomes: An Annual Report on Outcomes and Assessment Activities at Public Colleges and Universities in New Jersey (1989)

Report to the Board of Higher Education on the First Administration of the General Intellectual Skills (GIS) Assessment (1990)

Traipsing into Tricky Territory: Assessment of Student Personal Development, Involvement and Satisfaction Outcomes

A Report on Access, Retention, Transfer, and Graduation At New Jersey's Public Colleges and Universities (1991)

APPENDIX B

An Excerpt from The Final Report of the Student Learning Outcomes Subcommittee, 1987 (Pages 25-28) Providing an Operational Definition of General Intellectual Skills

I. Accumulating and Examining Information

a. Determine What Kinds of Information are Needed for a Given Task.

- Recognize when the necessary information is given e.g., in a specific reading assignment or in a lecture, and no further gathering of information is necessary.
- Recognize when additional information must be gathered, from a library (where the student "searches" for and "finds" information), from a laboratory (where the student "generated" information), or from other people (from whom the student "elicits" and "derives" information.
- Recognize when information must be extracted, e.g., from a work of art or literature, where the information must be gathered through a process of analysis.
- Determine which of these processes will be required to obtain the necessary information, and at what level of detail.
- Determine what kinds of information will be included and what kinds of information will be rejected.

b. Gather the Information Needed for a Given Task

- Construct an effective search procedure for gathering information on a given topic in a "library" -- reflecting an understanding of where to look, the various ways in which information is organized, and the various ways of accessing information, both computerized and non-computerized, and the parameters (e.g. category, key-word, etc.) of the information needed to develop a search strategy.
- Construct an effective search procedure for gathering information on a given topic in a "laboratory" -- reflecting an understanding of how data is gathered in a variety of settings, and by a variety of techniques.
- Construct an effective search procedure for gathering information on a given topic from other people -- reflecting an understanding of whom to seek out and how to ask appropriate questions.

c. Understand Information

- Absorb information, whether it is "given," must be "gathered," or must be "extracted."
- Replicate the information in a manner that accurately captures the original intent.
- Determine which information is needed for a particular task.
- Summarize the information that has been gathered, using notes which have been prepared and organized appropriately.
- Evaluate the information gathered in terms of relevance, credibility, importance, usefulness, and adequacy.
- Evaluate the information gathering process itself in light of the information that has been obtained. Determine whether additional information is needed and, if so, what more is needed and now it may be obtained.

II-A. Reconfiguring, Thinking About, and Drawing Conclusions From Information (Non-Quantitative)

a. Organize Information

- Evaluate the arguments and conclusions of others in terms of their assumptions, logical inferences, and the empirical evidence they offered to support their ideas, as well as one's own knowledge.
- Construct conceptual frameworks within which the information gathered can be organized in a way suitable to carrying out the task.
- Organize the information obtained in a variety of suitable ways within those frameworks and make judgements as to which ways are the most useful for carrying out the task.
- Select a conceptual framework and a way of organizing the information which appears most suitable, considering the information gathered and the purpose of the task.
- Reevaluate the information gathered as to adequacy, relevance, and usefulness within the framework chosen.

b. Think About and Draw Conclusions From Information

- Delineate a variety of interpretations, explications, or hypotheses which are compatible with the information.
- Evaluate the relative merits of those interpretations, explications, or hypotheses and select one or more which are worthy of further elaboration and testing.

- Delineate the logical implications of those interpretations, and draw conclusions from the analysis, using evidence contained in the information that has been gathered.

c. Evaluate the Results of This Process

- Determine whether the conclusions obtained were reasonable and whether the reasoning used to obtain those conclusions was sound.
- Evaluate whether the interpretations chosen for elaboration were appropriate, given the conclusions to which they have led.
- Reevaluate the choices of conceptual framework and organizing principles as to adequacy and usefulness in light of the conclusions obtained.
- Raise new and significant questions which are suggested by the conclusions obtained.
- Recognize when it is necessary or appropriate to repeat or to return to earlier steps in the process in light of this evaluation.

II-B. Reconfiguring, Thinking About, and Drawing Conclusions From Information (Quantitative)

a. Organize Quantitative Information

- Evaluate the interpretations of data done by others in terms of their assumptions, logical inferences, and the empirical evidence they used, as well as one's own knowledge.
- Construct one's own representations of a given situation, including, where appropriate, translations from verbal representations of the situation to arithmetical, algebraic, or statistical representations.
- Organize the existing information, within those representations, in ways suitable for the given task.
- Evaluate which representations appear to be most useful for carrying out the given task.
- Devise strategies or hypotheses for solving a given problem, and select one or more of those strategies for further elaboration.

b. Think About and Draw Conclusions From Quantitative Information

- Execute the arithmetic, algebraic, or statistical operations necessary to implement the representations or problem solving strategies selected or to test the hypotheses selected.

- Use those representations and problem-solving strategies to analyze and draw conclusions from the data.

- Display conclusions using the various ways in which quantitative information can be represented.

- Determine whether the data substantiate the hypotheses tested.

c. Evaluate the Results of This Process

- Evaluate whether the results obtained, and the conclusions drawn from those results, are plausible, and whether the reasoning used in drawing those conclusions is sound.

- Evaluate whether the representations, problem-solving strategies, and hypotheses are appropriate.

- Raise new and significant questions which are suggested by the conclusions obtained.

- Recognize when it is necessary or appropriate to repeat or to return to earlier steps in the process in light of this evaluation.

III. Presentation

a. Determine how one's results can best be presented and plan the stages of development necessary to achieve the desired end-product.

b. Carry out the various stages in preparing the material for presentation, including organizing the material, preparing an outline, displaying quantitative information appropriately, preparing a draft, and converting that draft into a finished product through various stages of revision.

c. Present the information that has been gathered and the conclusions drawn from that information in oral, written, and graphic formats, in ways that will be intelligible and persuasive to specified audiences and for specified purposes.

d. Evaluate, at each stage of preparing the material for presentation, whether additional information needs to be gathered or whether additional thought needs to be given to the framework, representation, or strategies selected, and to the conclusions that have been drawn from the given information.

APPENDIX C

DESCRIPTION OF SAMPLE TASKS

(Excerpt from the COEP Council Report, 1990)

Teresia and Conland: Students look for significant trends in various tables containing data about two fictitious countries, interpret the data, speculate on reasons for the trends in the data, and then compare the two, basically in terms of their economies.

Lemon Sharks: Students receive a map showing the breeding grounds of a large fish with data concerning its feeding, growth, migratory habits, territoriality, etc. They are then asked to trace the fish's movement and calculate population growth, extrapolating from birth rates and other information. They then receive information about sturgeon and are asked to state the ways in which the purposes and methods of studying the two fish might be similar or different.

Facts: Students receive a list of facts about a country's demographic makeup, the education of its citizens, the consumption patterns of its people, and the attitudes and values shared by the people. Students are then asked to identify relationships among these facts and draw conclusions about the society or culture of the country. Students indicate what additional information would be needed to support their hypotheses about this country's inhabitants.

Theories of the Universe: Students are given information about critical developments within the field of astronomy over a several hundred year period that shook the cultures of that time. Students are asked to evaluate the competing theories based upon their scientific merit, and then to account for peoples' reactions to these various theories emanating from the beliefs and values that were current at the time these events were taking place. Students then comment on the effect these changes had on humanity's sense of its place in the universe.

Sorting: Students receive general information about some of the geological processes that shape the world in which we live, and specific information about the layers of sediment formed at the bottom of a fictitious lake. They are asked to draw upon the information given to account for the pattern of sedimentation on the lake bottom.

Indo-Europeans: Students are asked to be historical linguistics in this task. They receive information about language families and their origins, and about the importance of a languages's core vocabulary in providing clues to the everyday lives of the speakers of that language. Students examine words from the core vocabulary of the original Indo-Europeans, as well as additional clues, and attempt to describe the group's original homeland and lifestyle.

REFERENCES

Board of Higher Education, A Resolution Establishing a Comprehensive Statewide Assessment Program; Trenton, NJ: Department of Higher Education, 1985

College Outcomes Evaluation Program (COEP) Advisory Committee, Report to the Board of Higher Education; Trenton, NJ: Department of Higher Education, 1987

College Outcomes Evaluation Program (COEP) Advisory Committee, Appendices to the Report to the Board of Higher Education; Trenton, NJ: Department of Higher Education, 1987

College Outcomes Evaluation Program (COEP) Council, Report to the Board of Higher Education on the First Administration of the General Intellectual Skills (GIS) Assessment; Trenton, NJ: Department of Higher Education, 1990

Educational Testing Services, A Report on the Development of the General Intellectual Skills Assessment; Princeton, NJ: Educational Testing Services, 1989

Morante, Edward A., The Effectiveness of Developmental Programs: A Two-Year Follow-up Study; Journal of Developmental Education, (3), 14-15, 1986

New Jersey Basic Skills Council, Report to the Board of Higher Education on the Results of the New Jersey College Basic Skills Placement Testing, Fall 1990 Entering Freshmen; Trenton, NJ: Department of Higher Education, 1991

Review of Edward Morante, "General Intellectual Skills (GIS) Assessment in New Jersey"

Reviewed by Richard L. Larson, Lehman College of The University of New York

This essay, as is again obvious, is less of a proposal or a suggestion about the conducting of a national assessment of students' attainment of our goal than it is a narrative of a series of events in New Jersey, an account of an assessment procedure (two procedures, really, the second built upon the first), an evaluation of that assessment procedure and a vindication of that procedure, and a declaration of the inherently political character of assessment. The lesson we are to take from Morante's narrative, I infer, is that regardless of how thoughtfully, painstakingly, informatively, and (as the author sees it) effectively an assessment is carried out, a shift in the political climate can, in effect, doom the effort and consign all that has been done, if not to oblivion, at least to an intellectual discard pile. It's not entirely clear from Morante's narrative whether the new state administration's treatment of the GIS assessment was due to the threat it posed to the reputations of educational institutions in New Jersey and/or to the possible progress of students, or simply to the arbitrariness (given Morante's narrative, one is tempted to say "rudeness") of that administration. Whatever the explanation, statewide assessment efforts, Morante seems to tell us, are fragile: the politics in control at any given time can support or destroy them, and can shift abruptly from support to destruction, regardless of how firm a consensus has been reached on the "skills needed for college education" and of how effective are the measures reached for assessing these skills.

Neither of the two assessment efforts that Morante reports deal directly, as I see them, with the kinds of assessment we are to undertake, or are carried on at the point in students' careers where we are to attempt our assessment. The basic skills assessment tests, as I understand them, were to enable the state to assure that students possessed the tools they needed in order to move successfully through college, though the test results might comment on students' high school preparation. And the GIS assessment dealt with students' accomplishments mid-way through college, not at the time of graduation. Still, Morante evidently sees the entire procedure as worthy of our attention, and takes particular pride in the plan developed for GIS assessment; he implies that his plan demonstrates the feasibility of at least a statewide assessment of general intellectual skills, and probably encourages the conducting of a nationwide assessment. I'm not convinced that Morante's perception of the possibility of moving from statewide to nationwide assessment is well supported; the procedures employed in New Jersey seem to have been notably

intricate, and the vast diversity of students and political/educational climates across the country might well, as Lanth says, present enormous obstacles to using one state's model nationwide. I think that the complexity of New Jersey's procedures would make them almost impossible to duplicate on a national scale, and I don't think that Morante tells us much of how such a national effort might work. Administering the individual tests so that not all students tested would have to deal with every question seems a reasonable idea, but the process of constructing the kinds of complex tests Morante talks about, and scoring on a very large scale the kinds of papers students produce on the tests, looks as if it would cause major administrative headaches, not to mention being very costly. The plan might work, but Morante tells us little of how it might work.

I have several unanswered questions about how the GIS testing was conducted in New Jersey. I'm eager to know more of how and why the testing procedures were judged valid and successful. I'm eager to know how the various exercises were developed and tried out. Most of all, I'm eager to know exactly how students' performances were judged, and by whom, and on what criteria, and I want to know more of how the criteria were developed and who trained the scorers to do the scoring. New Jersey may indeed have pioneered successfully in large-scale assessment of general intellectual skills, and have set in motion an effort that achieved the fifteen successes that Morante enumerates (pp. 10-11). But his essay is silent about a lot of the crucial issues involved in selecting a basis for judgment of students' performance, about the focus of readers' attention in the making of those judgments, and about the ways in which the judgments actually were made and validated. I applaud the effort to achieve judgments about individuals' verbal demonstration of the results of carrying on complex analytic, interpretive, and evaluative tasks--the tasks invented appear to have been ingenious in their demands for the cognitive/deliberative abilities to be assessed--and I don't right now envisage an alternative I could defend. But I want to know more of the details of what New Jersey did--details that are implicitly proposed as elements of a nationwide assessment procedure.

On one other point I find Morante's report valuable: its discussion of ways to assure that students invited for testing actually appear, and to assure that, having appeared, they do their best work. Morante gives absolutely appropriate attention to a problem any local or national testing project will confront, if it administers tests outside the context of students' work for a regular course. I wish I had had the chance to read his discussion of this matter before we at Lehman College undertook our assessment of the outcomes of our program in general education. If we had been informed by Morante's experiences, we could have conducted our assessment much more wisely.

PACIFIC GRADUATE SCHOOL OF PSYCHOLOGY

935 EAST MEADOW DRIVE, PALO ALTO, CALIFORNIA, 94303

(415) 494-7477

Evaluation Institute

Michael Scriven, Director

COMMENTS

1. EDWARD MORANTE ON THE NEW JERSEY TEST

This is an extremely valuable report on an extremely enlightening experience and it solved many of the key problems in a very enlightened way. I'll skip the detailed enconiums however, since we're trying to build on past experience, and cut straight to the bottom line on the debit side—since we're going to try to do better we need to know what needs to be improved.

First, the report does not provide us with the main reasons or rationalizations given for the canceling of the state-wide test, which would have been useful. The author may have felt diffident about doing so, since he was executive director of the project and its predecessor, and dismissed when it was terminated by a new governor's appointee, but I hope he will annotate his account informally with a word or two on this subject, from which further lessons might be learnt. Second, there were weaknesses in the test developed in NJ, many of them deliberately accepted as compromises in order to obtain some attractive trade-offs, include the following, which we should keep in mind:

• It can hardly be used to measure the achievements of individuals, since it would take 7 x 45 minutes i.e., 5.25 hours to administer, which for most of us rules it out for individual assessment. Using tests for program evaluation which have not been used

for individual evaluation involves some serious hazards about generalizability of the results; e.g., we don't know whether the profile of the idealized average student that results from the matrix sampling approach is a helpful profile for providing feedback to individuals. Since each student only does one question, a second limitation is that we can't find out anything about interaction effects.

• Since it uses essay response format, it is very expensive and time-consuming to mark.

• It is massively subject-matter loaded in that all of the items mentioned represent tasks which are common in certain subjects, rare elsewhere, and would provide an unfair advantage to students with experience in those subjects.

• There are serious validity problems, arising from various sources. The three most important sources are: (i) shared bias in that the scoring guides were developed mainly by people in the discipline to which the task mainly refers. If, for example, the task that concerns style in art history, does not have objective standards although those in the field think it does, the scoring guide will pass what we might call internal validity tests (those using judges in the field) but fail on external validity tests, i.e., those requiring the agreement of logicians assessing the field's own standards. This will lead to serious errors in evaluating the work of e.g., engineering students who share (in my view correctly) a high degree of scepticism about the objectivity of aesthetic standards. (ii) The second problem is coverage. The committee assumed, wrongly, that the test should mimic classroom tasks. This means that many non-academic uses of CT/PS were omitted, as also were subjects of importance which happen not to have achieved academic status yet. Use of established-subject-matter-connected material is politically attractive when you're trying to get acceptance by a committee of established-subject-matter-connected faculty, but it leaves out a vast range of skills which most non-academic people correctly perceive as part of critical thinking skills, e.g., the ability to analyze advertising and propaganda, appeals to emotion in editorials and graphic presentations (i.e., film or sheet advertising), practical product and service evaluation (e.g., the evaluation of teaching)—plus technology studies, crime detection, automobile and computer troubleshooting, home energy use, and other

subjects which have not yet received the accolade of curricular enshrinement—including the subject matter of CT itself—an understanding of terms like bias and objectivity and premise and assumption (there are some 70 of these logical terms in the standard English vocabulary). (iii) There are also weaknesses in the examples given with respect to the attitudinal component of CT and PS; for example, none of the examples provide a solid front of authorities agreeing on something which you are to criticize. This is one of the critical tests of CT skills, and one which sharply segregates even honors students, many of whom are simply incapable of the task of questioning authority.

(iii) The third validity problem is the contamination of the scores by essay-writing performance. While expression is rightly seen as one of the key higher-order skills, the facile writer tends to garner too many points for it at the expense of the logically crucial matters, as we confirmed in the Carnegie evaluation of the National Writing Project. Getting out of the frying pan of multiple choice items does not require getting into the fire of essays. There is an intermediate path, the use of what I have called Multiple Rating Items, which bypass the usual flaws in multiple choice items but retain fast scoring properties. A simple example of a multiple rating item requires the testee to allocate a grade to each of a set of answers to a common question; any of the answers can in principle get any grade so the 'lesser evil' algorithm doesn't work. More complex examples provide a repertoire of several critical or positive verbal comments, so that a more elaborate response can be constructed by checking the letters for more than one response, e.g., a grade and a reason for it. In either case, we are of course dealing with higher level skills, evaluation, synthesis, etc. and not recognitions skills. Of course the grading scale is defined, some of the items provide possible reasons for the grade, occasional write-in can be allowed. One can also allocate half marks for grades that are adjacent to the correct grade and alter the marking rubric so as to penalize answers which show a total lack of understanding, etc. A fuller account of this kind of item is provided in an appendix: "Multiple-Rating Items".

<u>Conclusion</u> The NJ experience involved the use of most of the state of the art testing

procedures, as seems appropriate in the light of ETS' location in that state; it's just that the state of the art in testing with respect to CT is not up with the state of the art in the field of informal logic. The gap is not vast, and it can be closed. However, it can't be closed for individual testing without using something other than essay tests and without rethinking and extending the domain tested.

POSITION PAPER EVALUATION
STUDY DESIGN WORKSHOP
HIGHER ORDER THINKING AND COMMUNICATION SKILLS
NATIONAL CENTER FOR EDUCATIONAL STATISTICS
17 - 19 NOVEMBER, 1991

Paper Reviewed: General Intellectual Skills (GIS) Assessment in New Jersey
by Edward A. Morante
College of the Desert

Reviewer: Ronald G. Swanson
Associate Director, Texas Academic Skills Program
Texas Higher Education Coordinating Board

This paper had a lot to say about the assessment of student thinking skills and could have applicability to a national assessment program as delineated in the National Goals for Education. New Jersey has struggled with the heavy issues associated with large scale measurement of college level skills and the lessons learned from that struggle should not be ignored. It was encouraging to see that performance assessment for large numbers of students is not only feasible, but actually workable in a practical sense. Therefore, there is much to be garnered from the New Jersey experience where the assessment of higher order thinking skills and communication skills are concerned.

It must be quickly added, however, that the New Jersey program leaves many questions unanswered where a national assessment program is concerned. Although most people would label the New Jersey program as "large scale", it is certainly not as large as those in other states and pales when viewed from a national perspective. Still, many of the principles and techniques used might be directly applicable to a national assessment design if the issues of funding, oversight and purpose can be more clearly defined.

In this review I will address certain aspects of the New Jersey program's strengths and weaknesses in light of possible use in a national assessment program. I ask the author's and readers indulgence since my comments are based solely on the information in the paper. There is a lot I'm sure I don't know about the New Jersey program and it will probably show.

1. The list of skills in Appendix B is quite good and includes many entries that should probably be in such a list for national purposes. As a matter of fact, the skill and subskill breakouts for the New Jersey program seem to fit the fifth objective of the fifth national goal quite well. However, the national goal specifically states that "The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially". Are the New Jersey skills representative of what a college graduate, as opposed to a freshman or sophomore, ought to be capable of? Are the skills listed advanced enough for college graduates, which seems to be the concern of the national goals? Of course, all of this assumes that at some point someone will

operationalize what is meant by the term "substantial increase" in objective five of national goal five.

2. The New Jersey paradigm is focused primarily on college students, which makes good initial sense. However, is there a developmental aspect to critical thinking, problem solving and effective communications? What if we find that college is too late to significantly influence those skills in our students? I refer here to the previously made point regarding the purpose of assessing the skills in question. Are we to simply assess and report the results - or will we use those results to cause some change in the national educational system? The national goal and the related objective lead me to surmise that in order to substantially increase the proportion of college graduates who demonstrate the skills in question, assessment would have to take place at several educational levels to determine when such skills are developed so that timely corrective action could be applied in order to affect the outcome relevant to the national goal. This may involve assessment in high school or even before. The core issue here is that we may have to assess prior to college in order to have any realistic chance of meeting objective five of goal five.

3. The New Jersey program assesses students while they are college sophomores. The national goal/objective specifies that the proportion of college graduates possessing the requisite skills is to increase substantially. There are two or more full years of college between college sophomores and college graduates. If sophomores have the appropriate skills - fine. If they do not, what is to be done? Are colleges and universities to offer special courses designed to remediate such deficiencies? Since the New Jersey model uses student sampling, an individual student will never know if he/she actually has command of the requisite skills or not. This drives us back to the likelihood that determining the "substantial increase" may well result from an assessment of the skills in question just before or just after college graduation with no time left to influence the result in college. This also implies that any educational changes that might be needed would have to be the result of feedback provided to earlier educational entities and efforts. The effectiveness of this approach does not appear to be good, at least where New Jersey is concerned, since high school graduate performance has changed little over the years as stated in the paper. On a national level, who would be responsible for bringing the "substantial increase" about? This has become a highly emotional issue in several states with colleges and universities resenting the remediation they must provide to students who graduated from high school without having the required skills. Any national program would have to step-up to these issues. The purpose of the assessment would have to be clear - would it simply be a barometer of current student performance or would it be used to drive changes to the educational system?

4. Several of the COEP variables given in the paper are affective in nature, i.e., "reasoned ethical responses" and "students' personal development values". While no one would argue that there are strong affective components

in thinking, communicating and learning, measuring them is quite another issue. Even if measured, what would be done with the results?

5. The New Jersey sophomore test emphasizes institutional responsibility rather than placing the burden of improvement on the students. As the author so aptly points out, this emphasis along with the attendant student sampling design doesn't demand much motivation on the part of individual students. This same phenomenon would be manifest in a similar national program. What would be the motivation to do well - especially on a national assessment? Then there is always the danger posed by volunteers and how they might be different from other students in systematic ways. The implications of this for a national assessment program are far reaching since any national program would, of necessity, almost be forced to use a sampling design rather than assess all students. While I agree with the author that the biggest problem would be getting students to come to the test itself, the assumption that most students will make at least some effort in completing the test once they are there is optimistic at best. Using student feedback in the form of test results as a motivator might very well work, but how could this be accomplished on a national scale? There was also no empirical evidence that I could find in the paper comparing motivated test results with unmotivated results - rather, it appeared that motivation was used primarily to get students just to show up.

6. The New Jersey program used tasks instead of multiple-choice items and kept costs down by using representative student sampling. As pointed out in the paper, the scoring process for such an assessment system is difficult, time-consuming and expensive. While the New Jersey system seems to have worked, maybe even worked well, what are the implications for a national assessment program? First, a national program would almost certainly have to use sampling. Even so, who would actually do the scoring, what funds would be available to pay for this, where would the scoring be done and to what criteria? New Jersey also developed extensive scoring guides and a scoring system. Who would develop such documents for a national effort? It might prove very difficult to get national agreement from all of the states as to what standards and criteria should be used. I would also like to resurface the possibility of using multiple-choice items for the national assessment effort. It is true that multiple-choice test items for the higher levels of learning are difficult to construct...but they can be constructed and are certainly easier and cheaper to score. This technique could result in being able to use larger samples with all the attendant strengths associated therewith. I would like to proffer that consideration be given to a national assessment that is not totally dependent on the more subjective task approach used in New Jersey - if for no other reasons than cost and practicality. If money, time and effort were not factors in a national program, the task approach, or something similar to it, might be the approach of choice.

7. The New Jersey program used a categorization system which classified student performance into three levels of proficiency: Demonstrated

Proficiency, Somewhat Proficient, and Did Not Demonstrate Proficiency. A system similar to this could be useful in a national effort. Proficiency levels would have to be carefully defined and agreed upon (that information was not given for New Jersey in the paper) and would probably better serve the national goal purposes than having a single score cut-off.

Again, much can be taken from the New Jersey experience. I fully agree with the author that "nationwide testing is possible not only at the level of basic skills but at a level of higher order skills as well". That is not in question - national assessment of higher order skills is possible. What is in question has more to do with another statement made by the author which asserts that "...such a concept can become operational given sufficient resources and leadership." Here will lie the crux of a national assessment program for critical thinking, communication skills and problem solving -- resources and leadership.